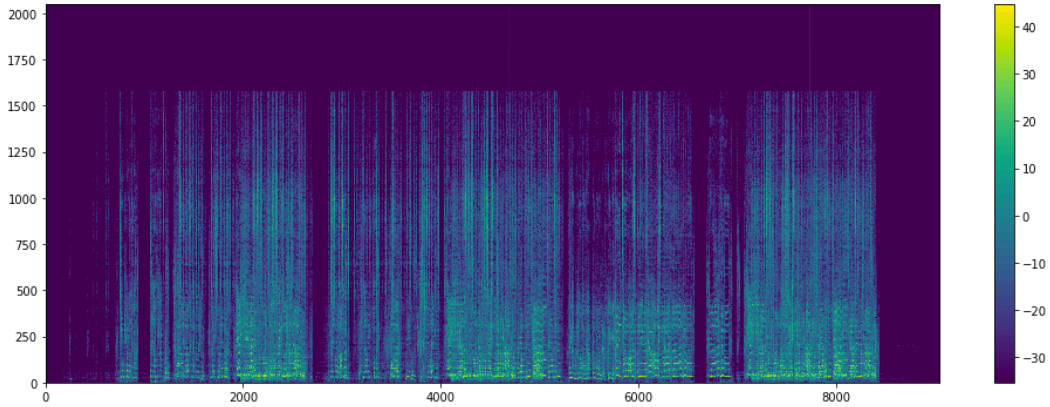
최종제안

LaSAFT 기반의 개선

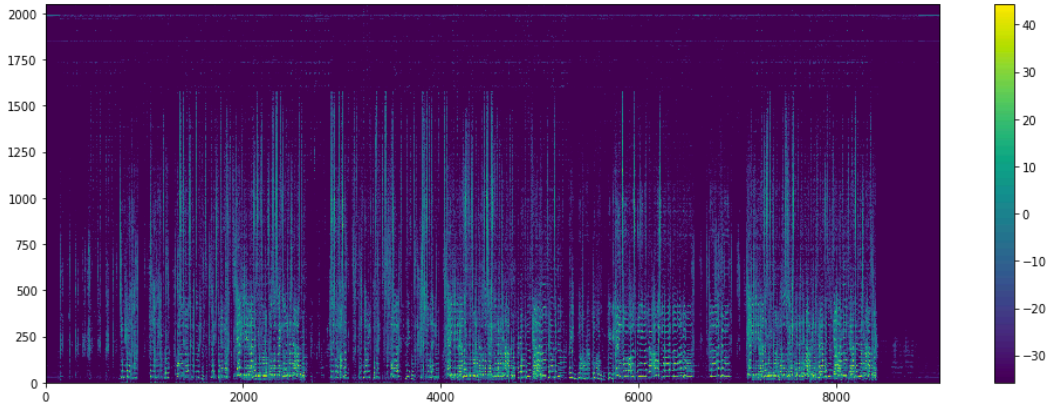
최종제안

Regression 방식의 문제점

Masking



Regression



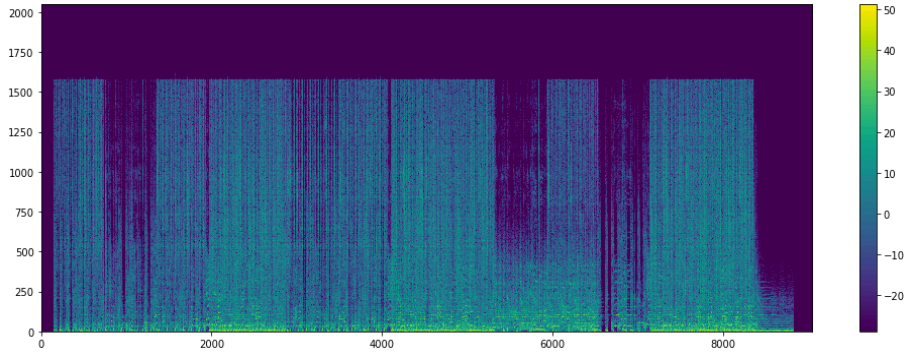
Masking	Regression
Noise decrease	Noise increase
드럼 분리 bad (특히 high hat)	드럼 분리 good
mixture phase 이용	***phase 복원** (성능향상에 큰 기여)

최종제안

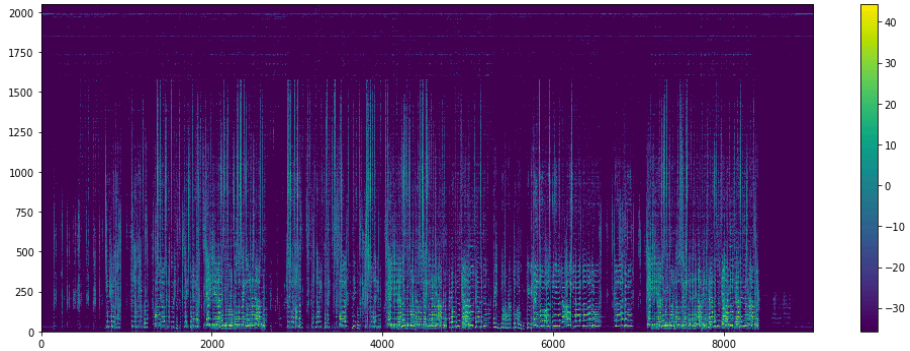
Regression 방식의 문제점

Mixture의 energy over 하는가?

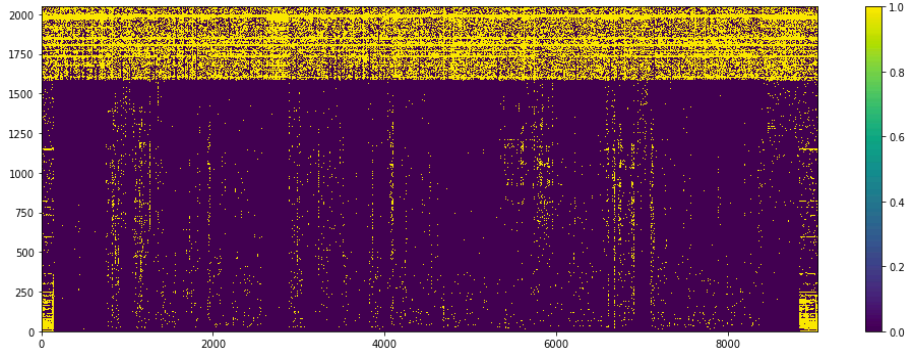
Mix



Predicted vocal



Beans over mix



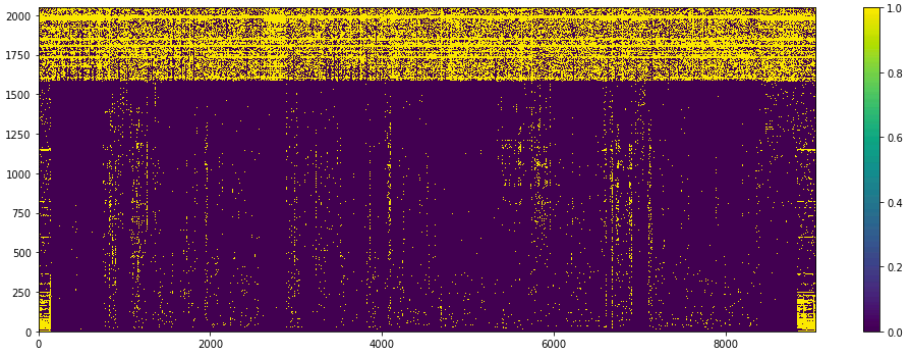
분리된 신호의 energy가 **mixture energy 넘는 경우**가 있음

최종제안

Regression 방식의 문제점

Mixture의 energy over 하는가?

Beans over mix



제안

$$loss = MSE(T, \hat{T}) + \lambda \sum max[0, \log(|\hat{S_{vocal}}| / |S_{mix}|)]$$

mixture 보다 energy 큰 경우에 penalty 주는 방식

최종제안

개선

제안1) **Short frame**에서의 **frequency pattern** 포착을 통한 성능 up

Replace **FC** to **1D Conv**

제안2) Noise 줄이기 위한 **regularized loss**

$$loss = MSE(T, \hat{T}) + \lambda \sum max[0, \log(|\hat{S}_{vocal}|/|S_{mix}|)]$$

mixture 보다 energy 큰 경우에 penalty 주는 방식

최종제안

개선

제안3) Loss for singing voice sep

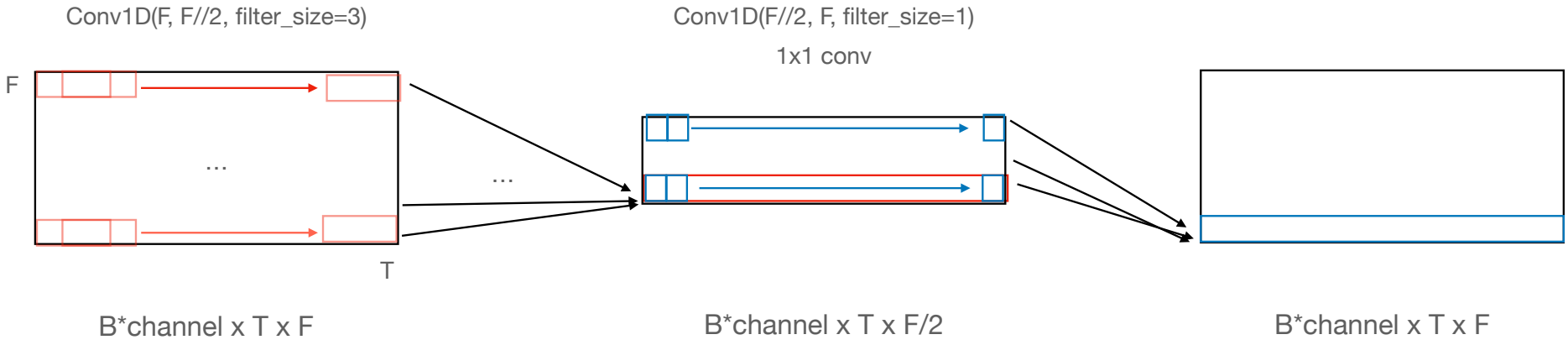
$$loss = MSE(T_{vocal} - \hat{T}_{vocal})$$

$$loss = MSE(T_{vocal} - \hat{T}_{vocal}) + MSE(T_{acc} - \hat{T}_{acc}) \ where \ \hat{T}_{acc} = T_{mix} - \hat{T}_{vocal}$$

하나의 모델에서 반주와 보컬 모두 잘 분리하게 끔 학습

최종제안

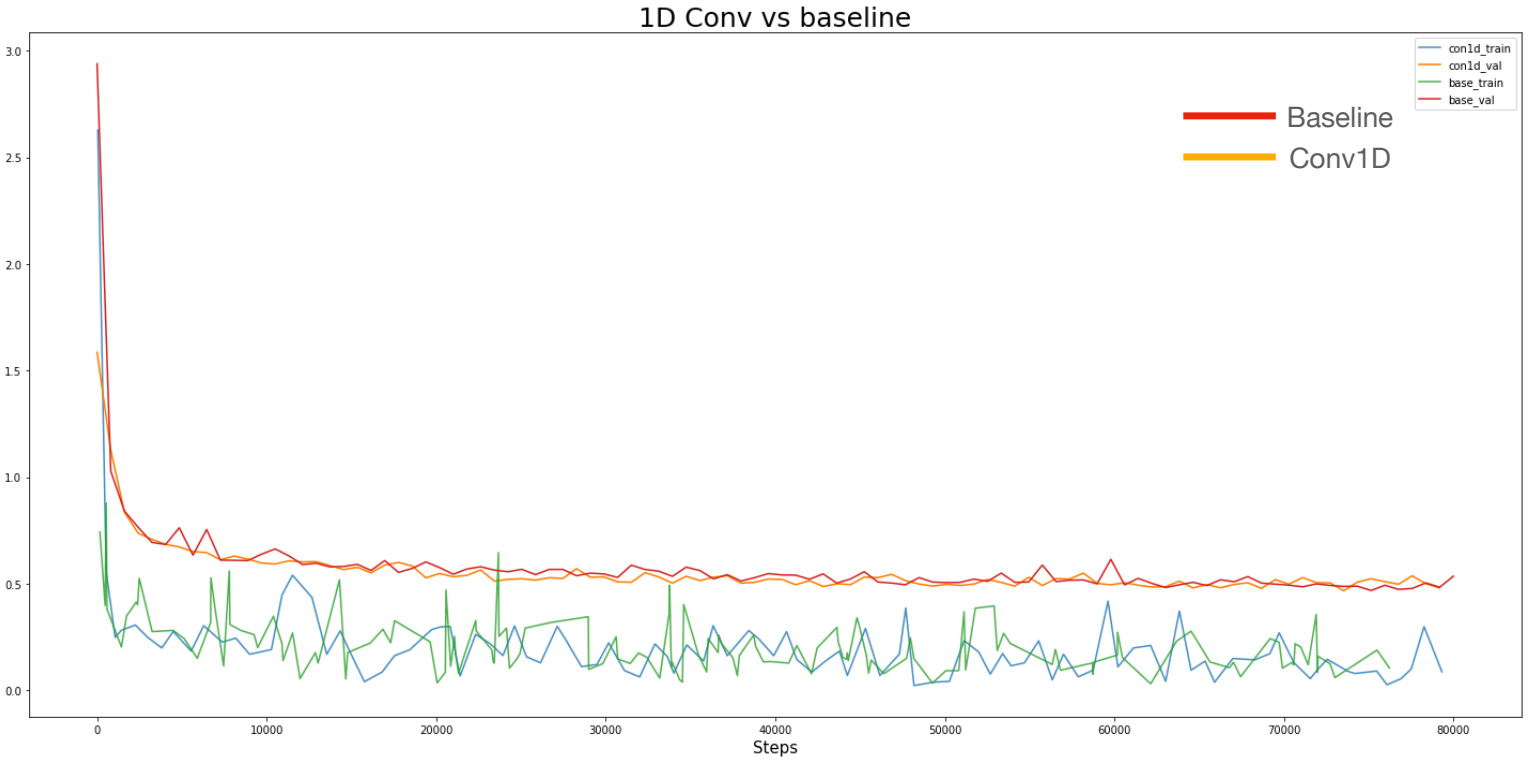
1D convolution



To learn common Frequency pattern

최종제안

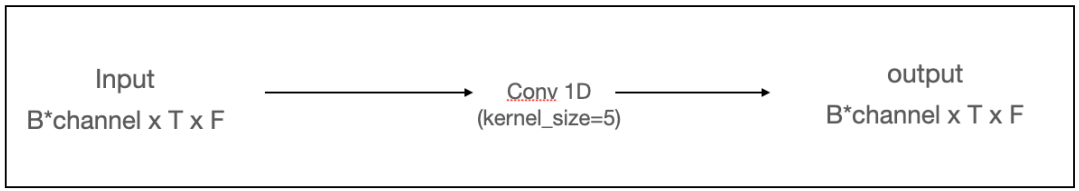
1D convolution



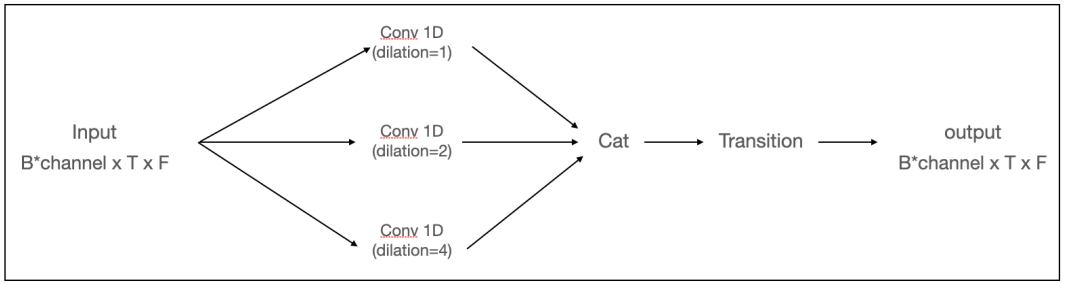
최종제안

1D convolution

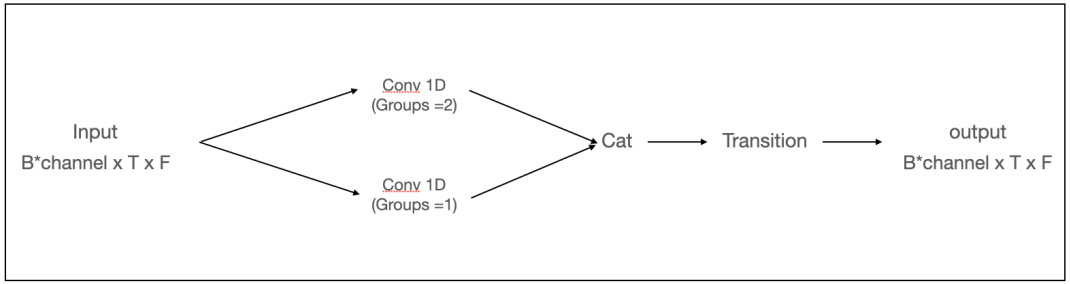
1D Conv block



Multi dilated Conv block



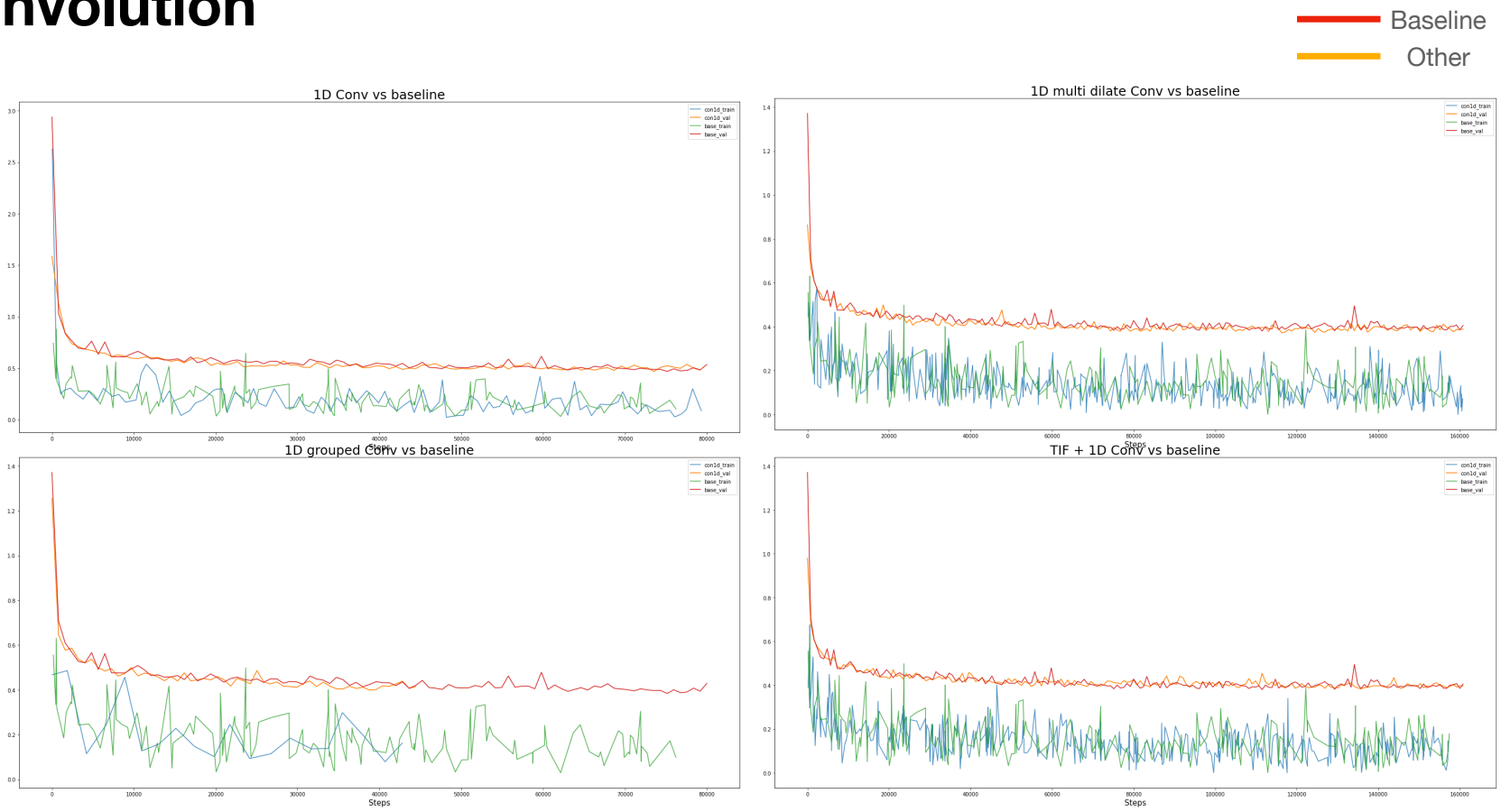
Band dedicated Conv block



4가지 방식으로 적용

최종제안

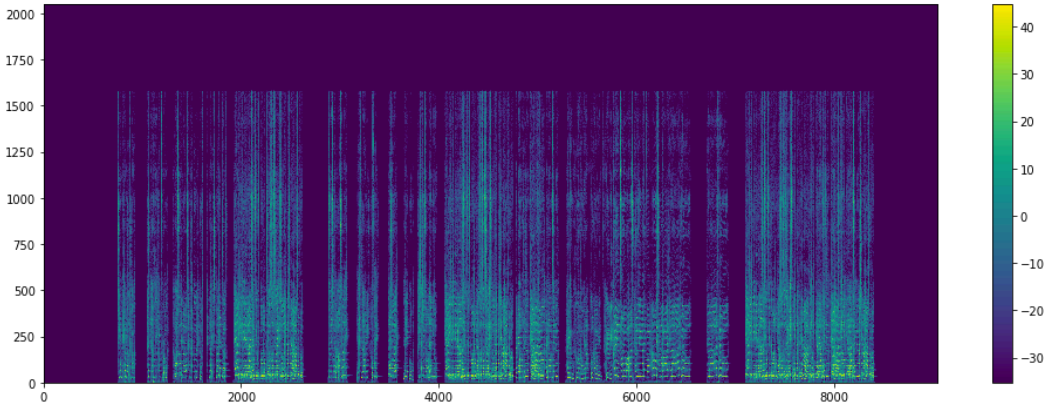
1D convolution



최종제안

1D convolution

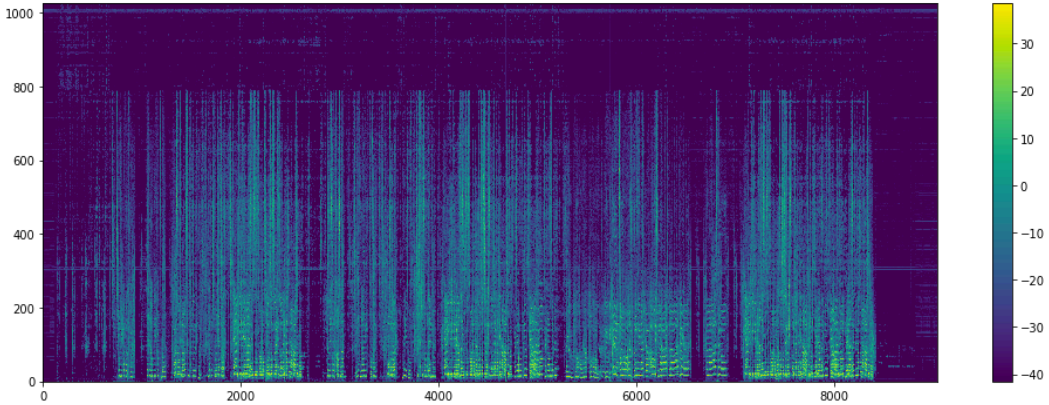
Reference



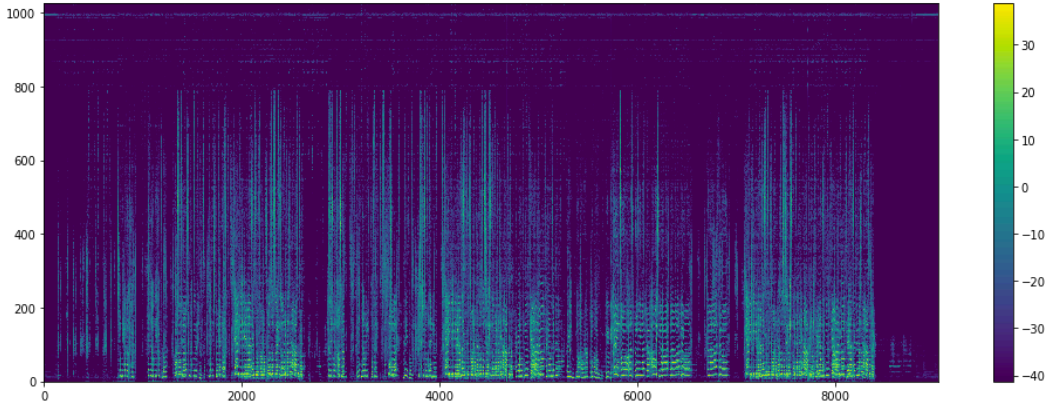
Conv의 경우, **high frequency 영역 많이 거른다.**
- **band dedicate mechanism** 적용

외각에서의 noise 많이 발생
- **regularized loss** 사용

FC predict (Baseline)



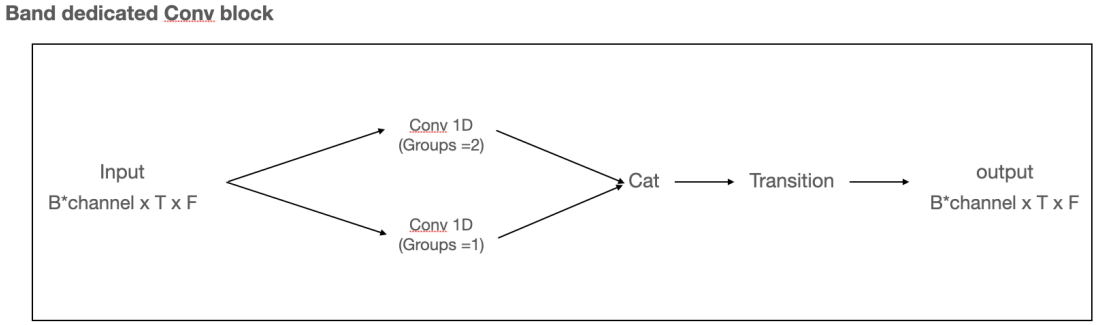
Conv predict



최종제안

1D grouped convolution + regularizer

1.



통한 **band dedicate** 형태

2.

By relu

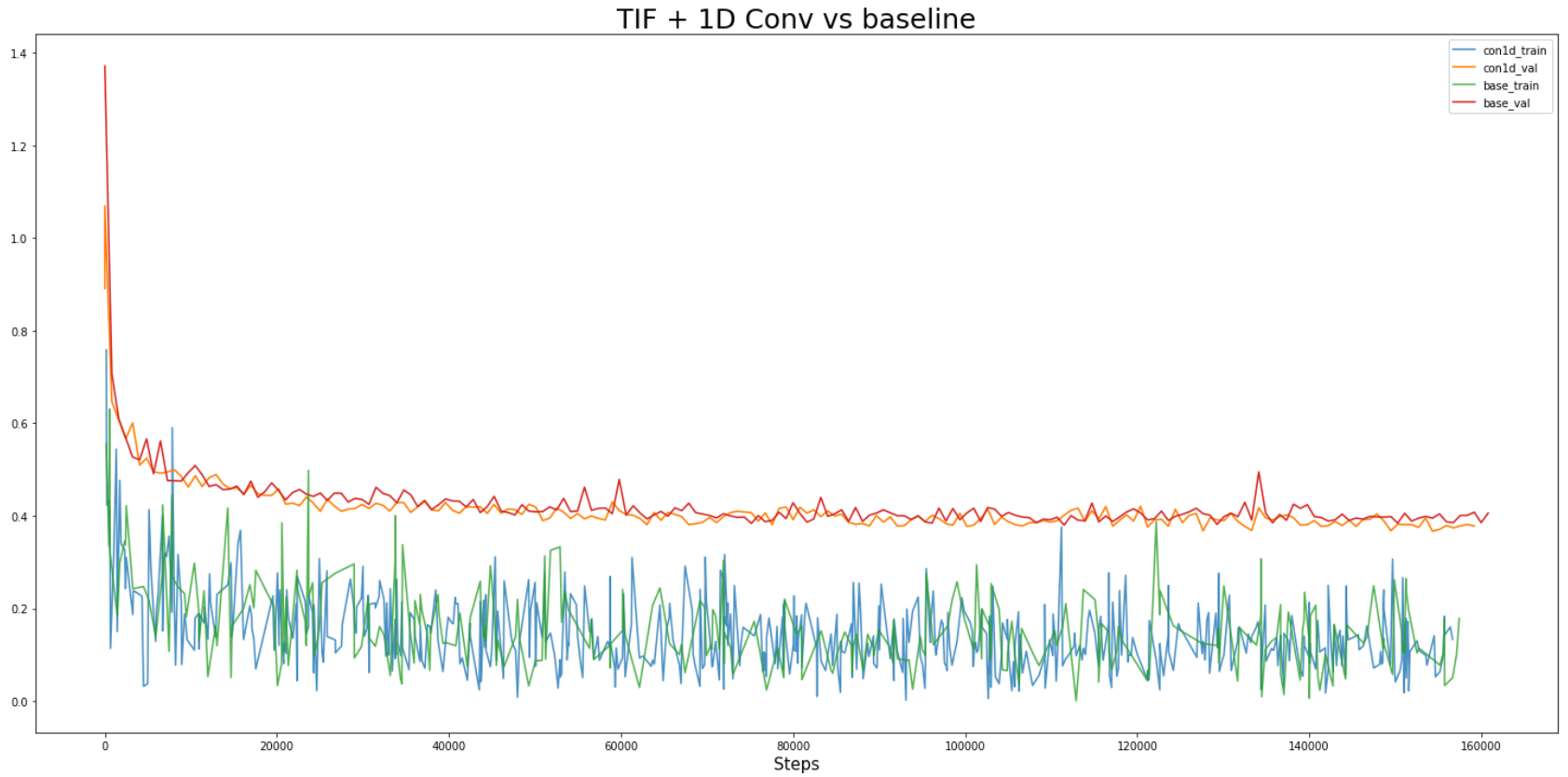
$$loss = MSE(T, \hat{T}) + \lambda \sum max[0, \log(|\hat{S}_{vocal}|/|S_{mix}|)]$$

추가 통해 노이즈 제거

mixture 보다 energy 큰 경우에 penalty 주는 방식

최종제안

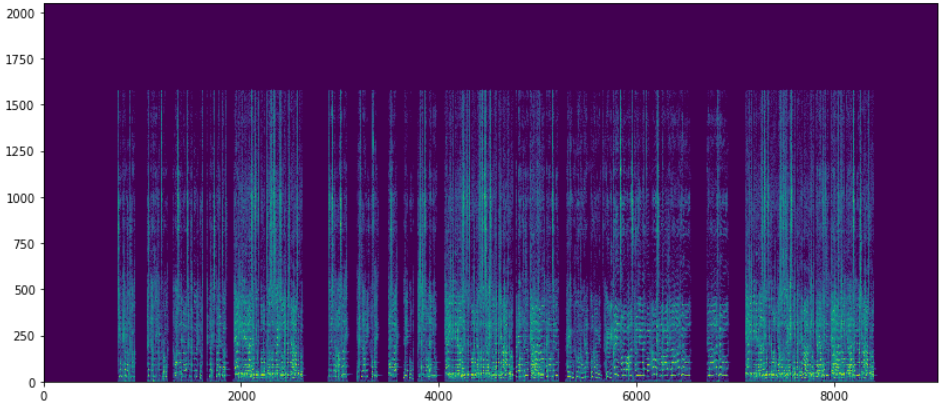
1D grouped convolution + regularizer



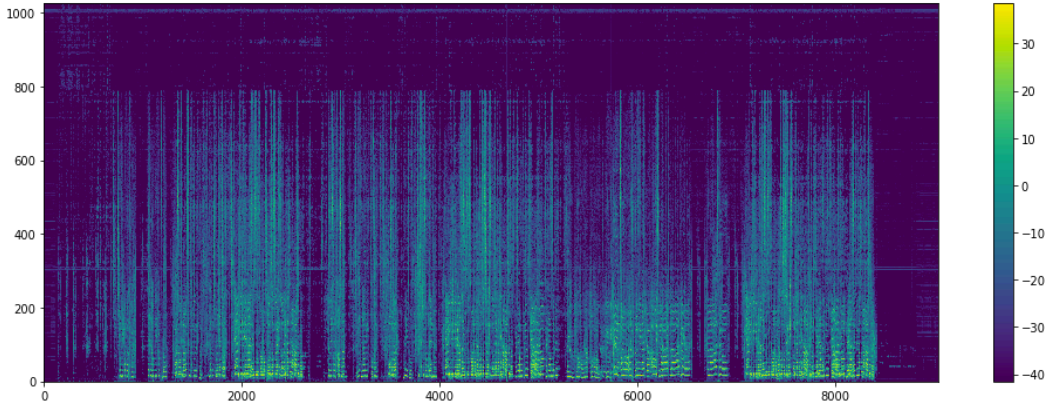
최종제안

1D grouped convolution + regularizer

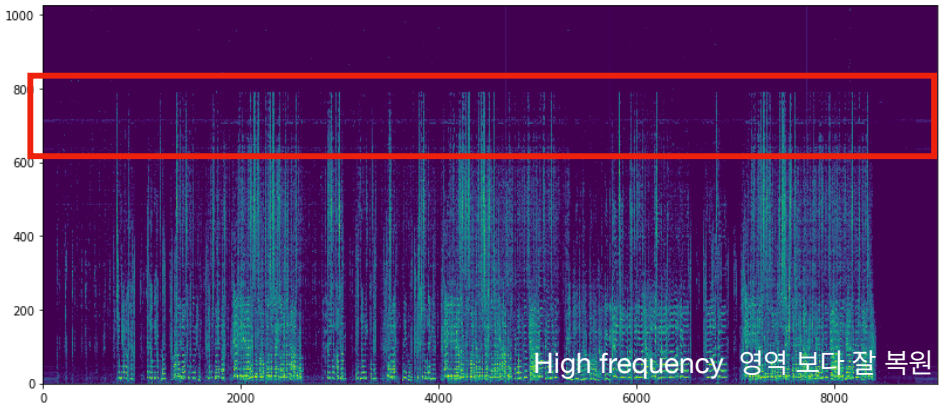
Reference



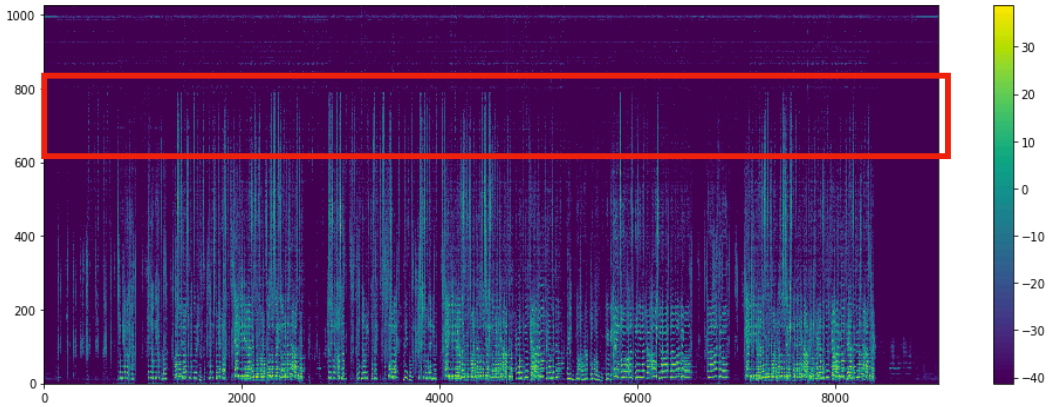
FC predict (Baseline)



Conv predict (Grouped + regularized)



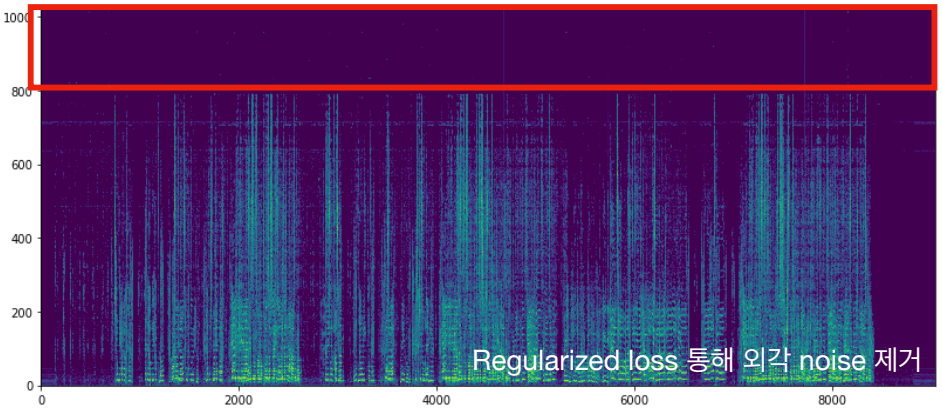
Conv predict



최종제안

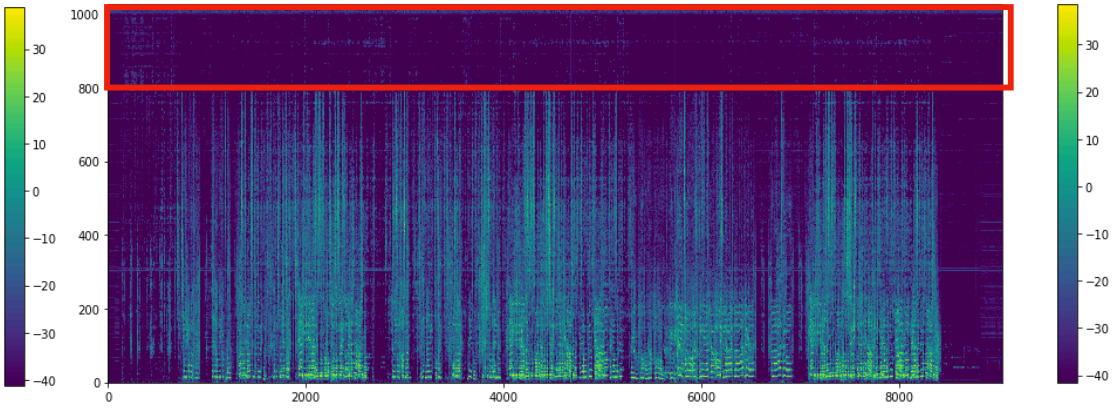
1D grouped convolution + regularizer

Conv predict (Grouped + regularized)



300 epoch
val_loss = 0.22075

FC predict (Baseline)

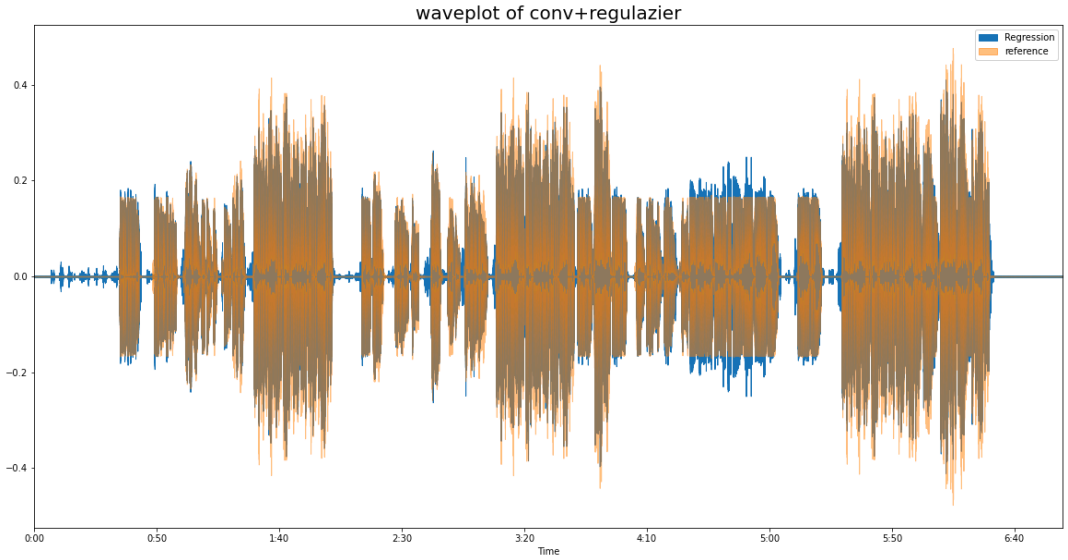


300 epoch
val_loss = 0.22506

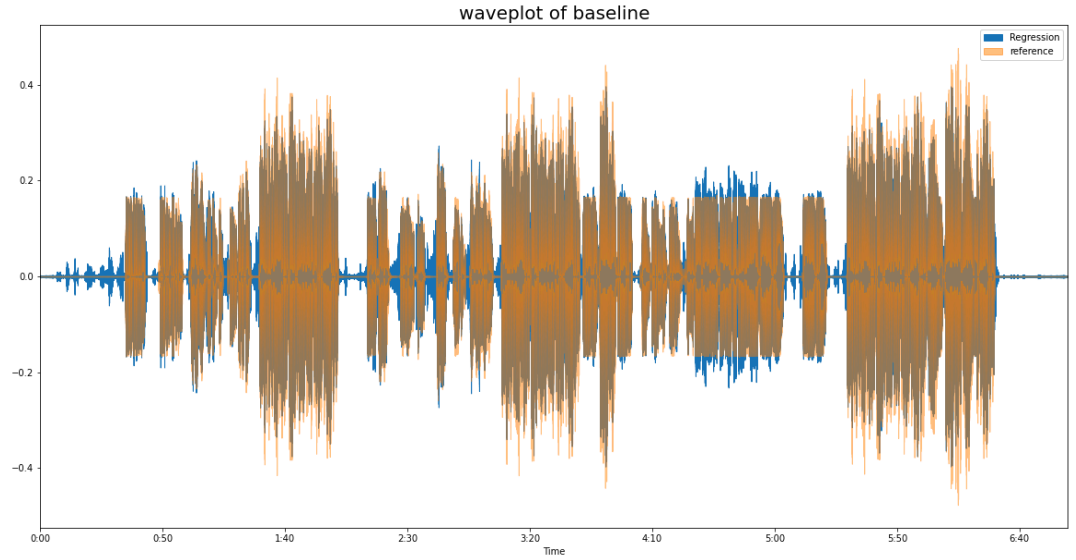
최종제안

1D grouped convolution + regularizer

Conv predict (Grouped + regularized)



FC predict (Baseline)



최종제안

1D grouped convolution + regularizer

Baseline

```
Aggrated Scores (median over frames, median over tracks)
vocals          ==> SDR:   7.191  SIR:  13.606  ISR:  12.850  SAR:   6.936
accompaniment   ==> SDR:  13.528  SIR:  18.111  ISR:  21.301  SAR:  14.611
```

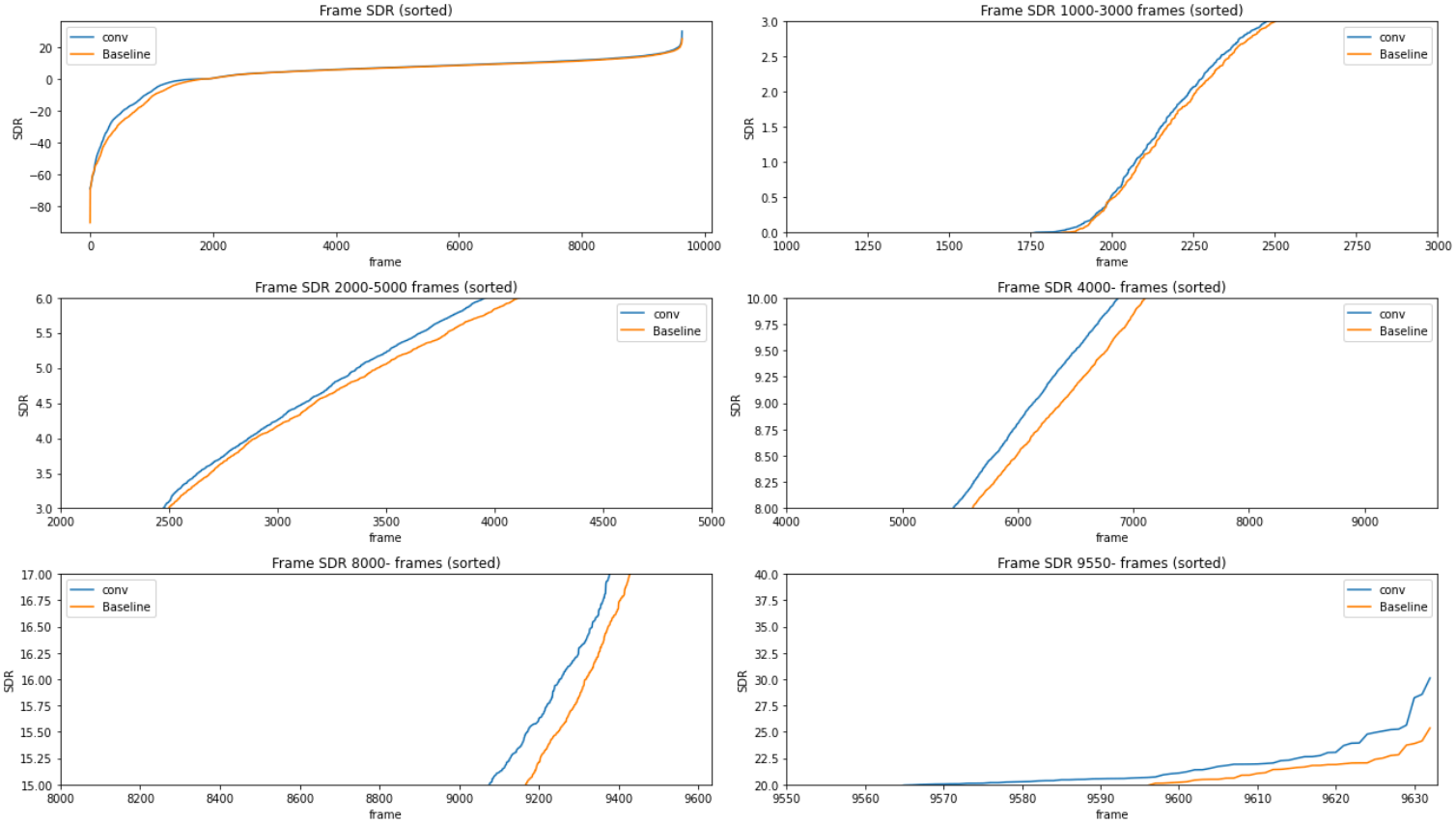
1D Conv + regularizer

```
Aggrated Scores (median over frames, median over tracks)
vocals          ==> SDR:   7.091  SIR:  16.144  ISR:  12.972  SAR:   7.054
accompaniment   ==> SDR:  14.082  SIR:  18.563  ISR:  22.853  SAR:  14.717
```

반주 SDR, 전반적인 수치 큰 폭 개선
but 보컬 SDR 소폭 하락으로 보임

최종제안

1D grouped convolution + regularizer



허나 **SDR** 역시 **전반적으로**
보았을 때 **더 좋아진것**을 알 수 있음

최종제안

결론

Grouped convolution + **regularized loss** 통해
큰 폭의 성능향상을 관찰

부록

부록

코드

Spleeter project

- CUNet
  - LaSAFT 관련 코드

- Sep_system
  - weight converter
  - Evaluation
  - Trainer
  - Models
  - Separator
  - Dataset
  - Preprocessing

- spleeter_2
  - Spleeter 관련 코드